DRAFT MEMO

Subject: Cross Validation of Two Methods for Event Cluster Analysis

From: Hans Israelsson (CMR), Bob Engdahl (CU) and Eric Bergman (CU)

Date: 2001/02/23

The Group 2 consortium will be using two published methodologies to construct station travel time corrections for clusters of well located events with and without ground truth information. Given similar or identical input data, we want to cross-validate the two methods to ensure that they produce consistent results - epicentral locations, error ellipses, and station travel time corrections. We also want to understand any systematic differences in results.

One of the methodologies employs the Hypocenter Decomposition theorem (Jordan and Sverdrup, 1981) and so called groomed ISC data (Engdahl et al, 1998). The other applies standard Joint Hypocenter Determination (Douglas, 1967) using an algorithm by (Dewey, 1972 and 1989) to ISC (2001) and NEIS (2001) data screened for outliers. Below we refer to these two methodologies as HDC and JHD respectively and compare results for event clusters near Azgir at the Caspian Sea (underground nuclear explosions), at Racha in Western Caucasus, and in the Gulf of Aqaba. The IASPEI91 travel time tables are used by both HDC and JHD, which both also use only first P arrival times. Detailed discussion of the two methodologies is beyond the scope of this memo. However, some of the relevant differences are discussed briefly.

In summary, the comparisons did not reveal any statistically significant differences for the purposes envisioned by the Group 2 consortium. Some minor systematic differences were noted in the size of error ellipses due to the weighting of input data and in application of ellipticity and elevation corrections. These systematics are worthy of some further investigation but do not impact the independent use of the two methodologies for estimating station corrections for clusters of well located events with and without ground truth information.

AZGIR

Seven closely spaced underground nuclear explosions near Azgir with ground truth information reported by Sultanov et al., (1999) (see Table 1 below) were used to test and compare the accuracy of the methods In the comparison for Azgir explosions, HDC and JHD were applied to the *same* arrival data to validate computational consistency of event locations and station travel time corrections. Identical arrival data also affords the opportunity to compare the scaling of error ellipses by the different data weighting of the two methods. Table 2 summarizes the HDC and JHD locations relative to the ground truth and gives associated error estimates. The resulting locations with associated error ellipses are also compared in Figure 1. The HDC and JHD epicenters are generally

consistent and have a maximum separation of 1.7 km. The HDC and JHD error ellipses also overlap ground truth within about 0.5 km.

Table 1: Ground Truth data for Azgir explosions (Sultanov et al, 1999)

Event No	Date	Time	Lat(N)	Lon(E)
1	1976/07/29	05:00:00.50	47.870	48.150
2	1977/09/30	06:59:58.43	47.897	48.161
3	1978/10/17	04:59:59.06	47.850	48.120
4	1978/12/18	07:59:58.50	47.860	48.160
5	1979/01/17	07:59:58.50	47.920	48.120
6	1979/07/14	04:59:58.00	47.880	48.120
7[a]	1979/10/24	05:59:59.00	47.850	48.140

a. Constrained in JHD locations

Because of differences in weighting of arrival data HDC and JHD error ellipses have slightly different orientations (median difference 16 degrees) and ellipticity (within a few per cent). The JHD ellipses are, on average, slightly larger with the semi axes being about 10% longer (median), which corresponds to the ellipse area being about 20% larger than that of HDC. The median of the origin time differences (HDC - JHD) was about 0.1 sec with a maximum difference of 0.2 sec.

Table 2: HDC and JHD locations relative to Ground Truth and error ellipses (90%)

Event No.	Δtime		Δlat		Δlon		Error Ellipses[a]					
	HDC	JHD	HDC	JHD	HDC	JHD	HDC			JHD		
							a	b	φ	a	b	φ
1	-0.46	-0.50	0.000	0.008	-0.015	-0.016	1.5	1.3	23	1.7	1.4	6
2	-0.30	-0.53	0.008	0.005	0.029	0.029	2.1	1.8	28	2.0	1.8	5
3	-0.35	-0.46	-0.002	-0.006	0.000	0.001	1.5	1.2	19	1.7	1.3	6
4	0.21	0.10	-0.017	-0.015	-0.002	0.015	1.5	1.2	22	1.7	1.3	6
5	-0.69	-0.80	-0.005	0.001	-0.001	0.001	1.7	1.3	13	1.8	1.4	7
6	-0.54	-0.60	0.004	0.009	0.002	0.004	1.5	1.2	21	1.7	1.3	5
7[b]	0.14		0.013		-0.012		1.5	1.2	22			

a. a = semi major, b = semi minor axes(km), φ = strike angle (degrees from north)
b. Constrained in JHD locations

Table 3 compares travel time corrections for stations within 20^o which are also plotted in Figure 2 and Figure 3. Corrections for station elevation and ellipticity were made for HDC but not for JHD travel times. The effect of this procedural difference is estimated to be less than 0.3 sec. The standard deviations (σ) of the station travel time corrections were also estimated somewhat differently (HDC uses normalized median absolute residual deviation while JHD estimates are weighted standard errors). The error bars in Figure 2 represent 90% confidence intervals for the estimated corrections, which indicate agreement across the entire range from -6 to 2 sec of the travel time corrections.

Table 3: Station travel time corrections (bias) for Azgir events

Station	Dist	HDC			JHD		
		Bias	σ	nobs	Bias	σ	nobs
BEO	19.3	-1.30	1.04	5	-1.18	0.99	5
DEV	17.4	-3.45	0.30	6	-3.63	0.29	6
ELL	17.4	1.90	0.59	5	2.00	0.87	5
ISK	15.2	-3.90	0.15	5	-3.74	0.25	5
JOS	18.4	-5.25	0.37	6	-4.99	0.47	6
KAS	12.1	-1.80	0.30	5	-1.69	0.21	5
KER	13.5	0.70	0.15	5	0.81	0.64	5
KJF	19.8	-4.30	0.44	7	-4.34	0.47	7
KRA	18.6	-5.90	0.44	7	-5.70	0.42	7
MAIO	14.3	-3.35	0.67	6	-3.06	0.43	6
NIE	18.4	-5.08	0.19	5	-4.80	0.75	5
NUR	18.6	-6.08	0.11	7	-5.93	0.10	7
SHI	18.5	-1.34	0.07	7	-0.78	0.09	7
SKO	19.7	0.09	0.06	7	0.18	0.18	7
SPC	18.5	-2.62	0.07	6	-2.21	0.41	6
TAB	9.9	1.95	0.30	6	2.12	0.23	6
TBZ	9.1	-0.10	0.89	5	-0.14	0.87	5
TEH	12.3	-0.80	0.89	5	-0.12	1.09	5

In Figure 3 the estimated travel time corrections are formally compared with confidence intervals of the differences (HDC-JHD) in travel time corrections, which are plotted as a function of epicentral distance. The confidence intervals for the differences are approximations of the case with two independent Gaussian samples having unknown means as well as standard deviations (Behrens-Fisher's problem, Cox and Hinkley, 1974). The confidence intervals cover the median offset

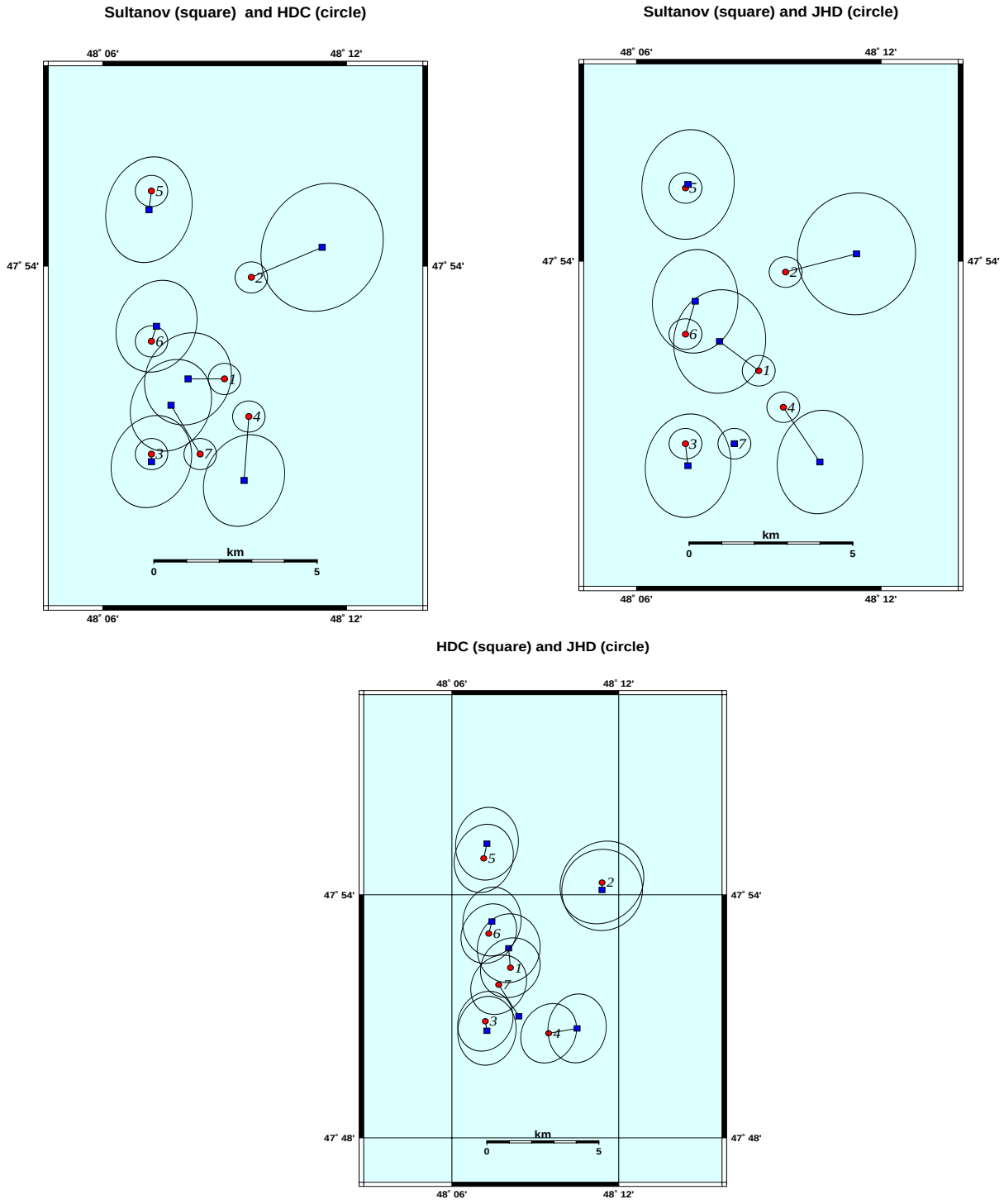

Figure 1. Comparison of HDC and JHD locations with ground truth epicenters by Sultanov et al. (1999) for Azgir events: Sultanov and HDC (upper left), Sultanov and JHD (upper right), and HDC and JHD (bottom).

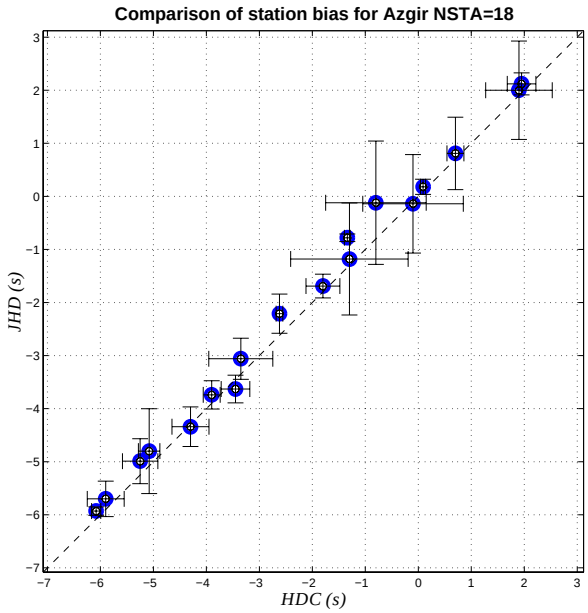

Figure 2. Travel time corrections for 18 stations at epicentral distances < 20^{o} obtained by HDC and JHD for Azgir events. Error bars indicate 90% confidence limits.

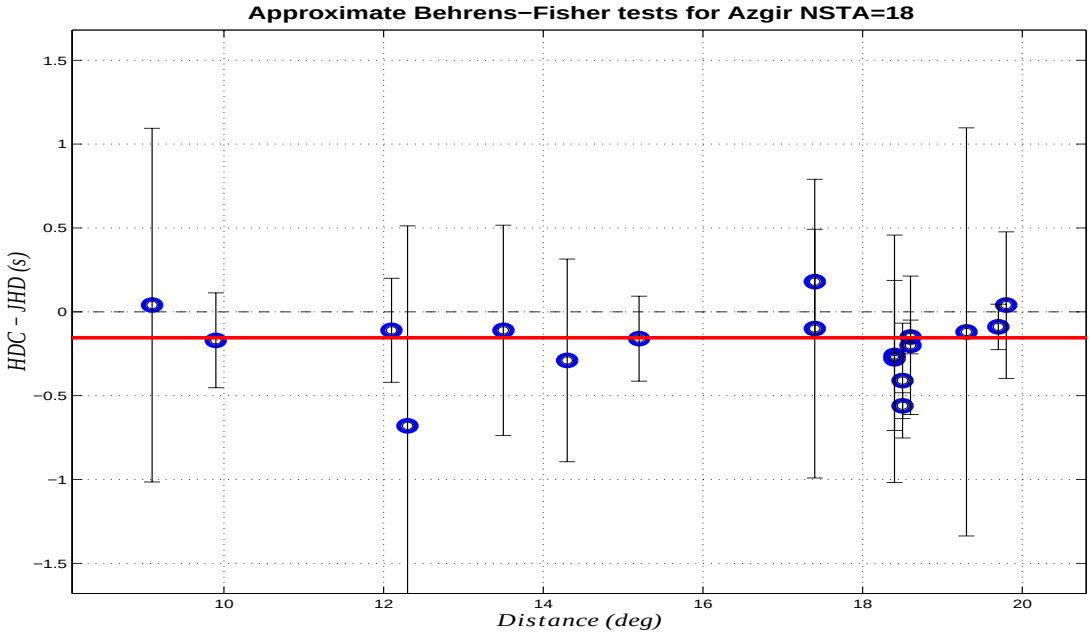

Figure 3. Difference in travel corrections plotted as a function of epicentral distance for Azgir events. Error bars represent approximate 90% confidence limits (Behrens-Fisher) for estimated differences relative to the median of the differences (-0.1 sec), which is marked as a thick line.

(-0.1 sec) for 16 of the 18 stations. The offset in station corrections can be directly related to the difference in origin times (see Table 2). The degree of overlap of the confidence intervals in Figure 3 is compatible with their 90% significance level. Scatter in the differences may be contributed by the fact that station elevation and ellipticity corrections were applied only for HDC estimates. Furthermore, for small samples with the number of observations being between 5-7, the HDC estimates of standard deviations (normalized median absolute deviation) may be biased low compared with the traditional JHD estimates.

RACHA, WESTERN CAUCASUS

HDC and JHD cluster results for the Racha sequence were obtained using different but overlapping data sets. The HDC results were based on 47 events that occurred between April-October 1991 and included a few reference events (ground truth category GT5) recorded by a temporary aftershock network. The JHD results were based on two clusters, referred to here as Racha East (24 events between April-1991 and Feb 1993) and Racha west (25 events between April, 1991 and Dec, 1994), according to their relative locations along the East-West direction of the whole Racha sequence. Solutions of events in the HDC cluster were used to constrain the JHD results, one for Racha East and one for Racha West. These two HDC solutions had been "promoted" to reference event of category GT5 by virtue of their semi-major error ellipse axes being less than 5.5 km. Table 5 lists HDC solutions, all of which had been "promoted to" category GT5 for events common to the three clusters.

There are 13 events of category GT5 common to the HDC and the JHD Racha East clusters and 5 GT5 events common to the HDC and the JHD West clusters. Events constrained for the JHD are indicated in Table 4 and comparisons of HDC and JHD locations are given in Figure 4. Reference events located by the temporary aftershock network are assigned error circles with a radius of 5 km in the figure, while estimated error ellipses are used for HDC GT5 solutions. There is complete overlap of error ellipses and circles of the common HDC and JHD events both for the JHD East and JHD West clusters.

Figure 5 compares travel time corrections for common stations at distances less than 20^{o}. Data are plotted only for stations with estimates based on 5 or more events for both the HDC and the JHD corrections. There is general agreement between the two types of estimates, with an offset of about 0.5 sec for the Racha West case. This offset can be related to estimated and constrained origin times; the origin times and travel time corrections of JHD solutions are tied to the origin time of the constraining event. The comparisons in Figure 6 of the formal approximate confidence intervals for the differences in the estimated travel time corrections show differences on the 90% significance level for 4 out of 30 stations for Racha East and 1 out of 17 stations for Racha West. This suggests that the observed differences are reasonably consistent with the statistical variations due to data and procedural differences.

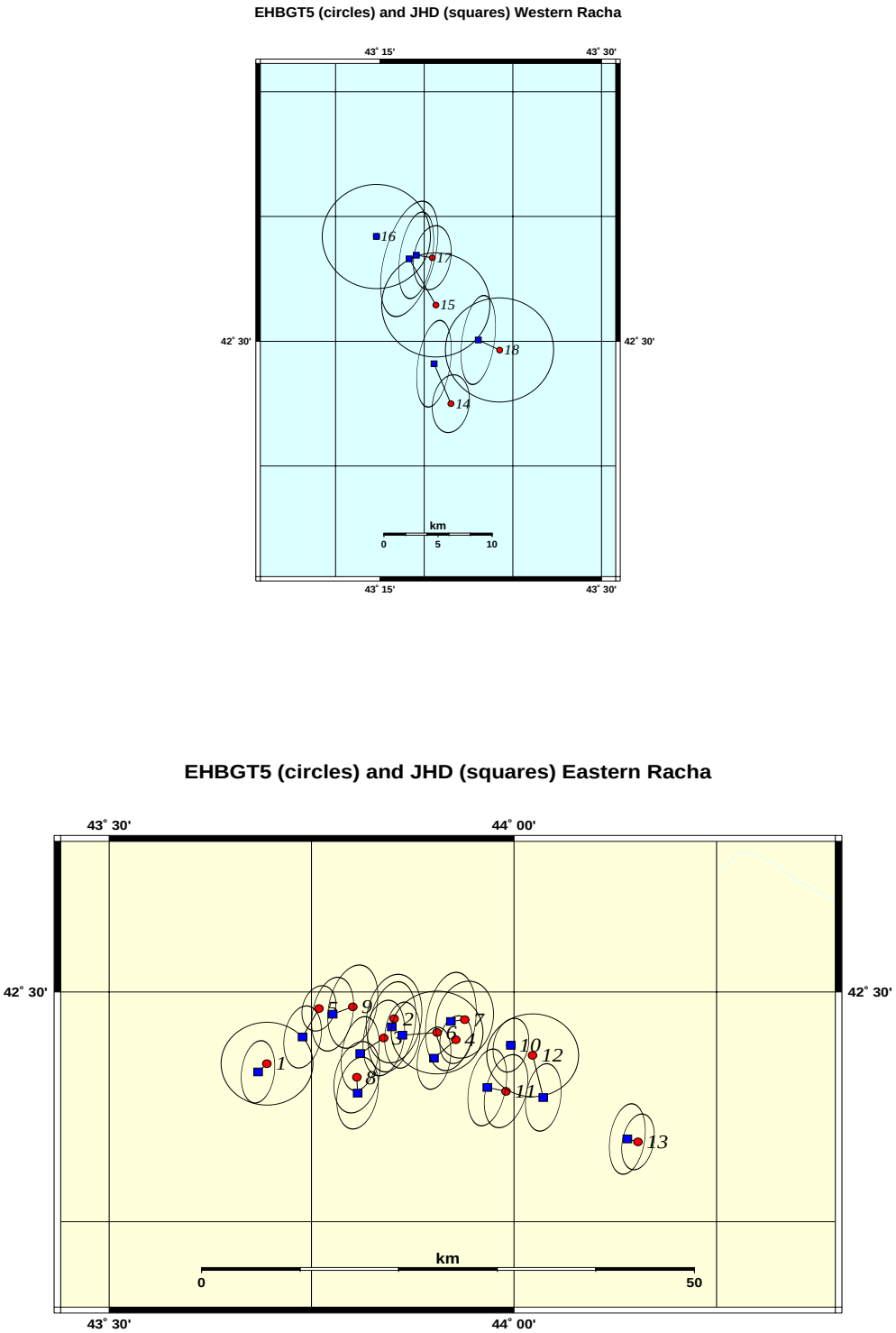

Figure 4. Comparison of JHD locations (squares) and reference events (circles) for western (upper frame) and eastern Racha event clusters (lower frame); event numbers refer to Table 4. Reference events located by temporary aftershock network are assigned an error circle with a 5 km radius, while HDC error ellipses are shown for other Racha reference events.

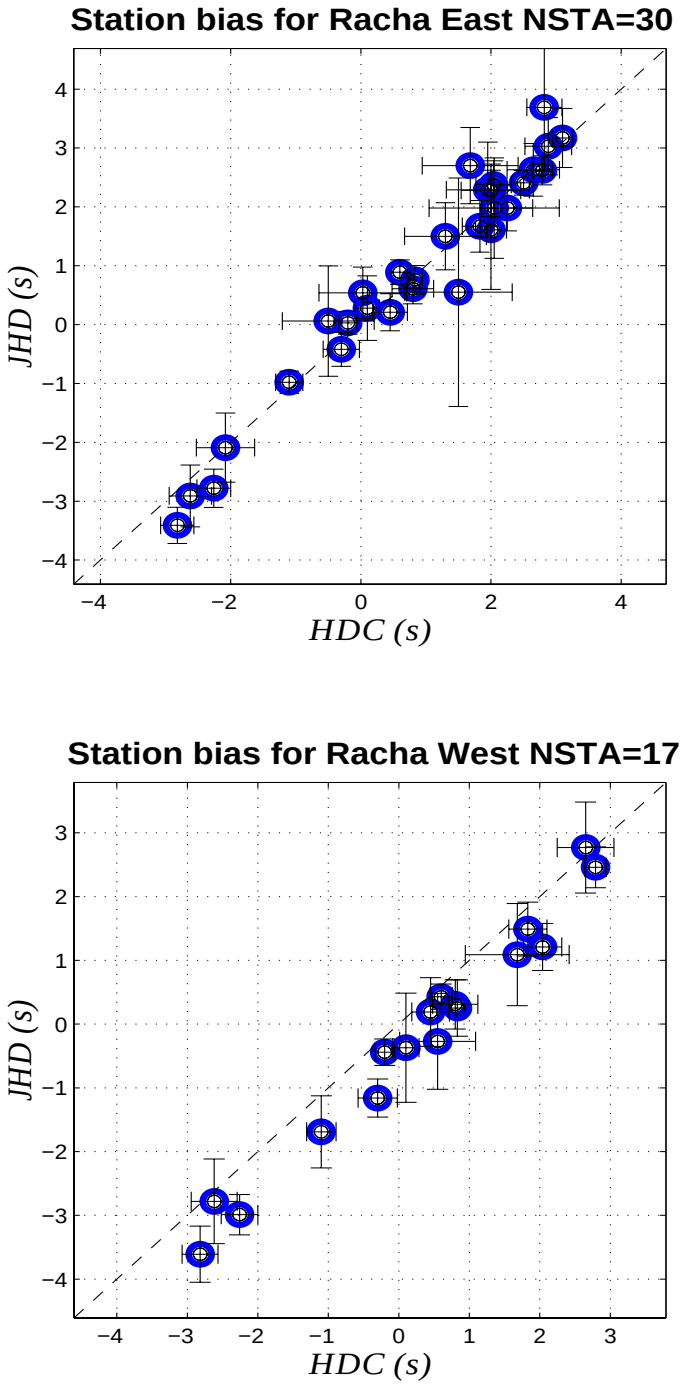

Figure 5. Travel time corrections for stations at epicentral distances < 20^{o} obtained by HDC and JHD for Racha eastern (top) and western cluster (bottom). Error bars indicate 90% confidence limits.

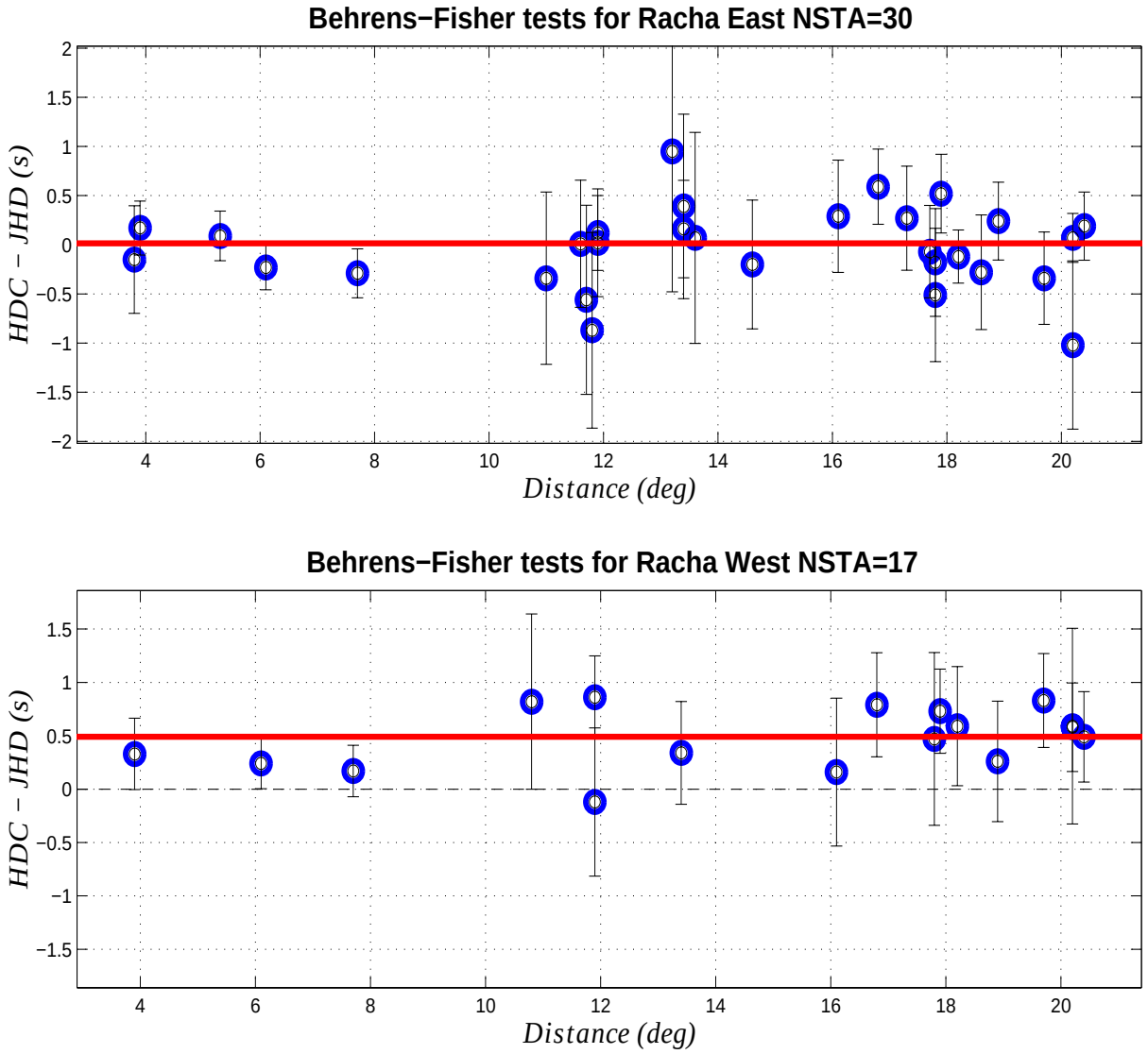

Figure 6. Difference in estimated travel corrections plotted as a function of epicentral distance for Racha eastern (top) and western event clusters bottom). Error bars represent approximate 90% confidence limits (Behrens-Fisher) for the differences relative to the median of the differences (0.0 sec for eastern Racha and 0.5 for western Racha), which is marked as a thick line in each frame.

.

Table 4: HDC GT5 solutions of reference events common to HDC and the JHD Racha clusters

Event No	JHD region	Date	Time	Lat(N)	Lon(E)
1	East	1991/04/29	09:12:46.40	42.4220	43.6950
2	East	1991/04/29	09:37:37.36	42.4710	43.8520
3	East	1991/04/29	11:51:10.27	42.4500	43.8390
4	East	1991/04/29	14:43:06.32	42.4480	43.9280
5	East	1991/04/29	18:23:15.12	42.4820	43.7590
6	East	1991/04/29	18:30:41.10	42.4560	43.9050
7	East	1991/04/29	19:44:54.45	42.4700	43.9390
8	East	1991/04/29	20:24:43.16	42.4070	43.8060
9	East	1991/04/29	22:28:22.70	42.4840	43.8010
10[a]	East	1991/05/02	01:25:29.73	42.4420	43.9960
11	East	1991/05/10	20:52:27.21	42.3920	43.9900
12	East	1991/06/15	00:59:20.10	42.4310	44.0230
13	East	1991/07/04	06:26:30.29	42.3370	44.1530
14	West	1991/04/29	20:32:54.19	42.4500	43.3300
15	West	1991/04/30	16:07:39.70	42.5290	43.3130
16[a]	West	1991/05/03	20:19:39.60	42.5840	43.2460
17	West	1991/05/03	23:41:01.84	42.5670	43.3090
18	West	1991/05/15	14:28:50.20	42.4930	43.3850

a. Constrained in JHD solutions

GULF OF AQABA

The comparisons for the Gulf of Aqaba were based on a single HDC cluster and three partly overlapping JHD clusters. The HDC cluster included 37 events distributed along the Gulf of Aqaba with a linear extent of about 100 km and the cluster was shifted using one reference event (data by Geophysical Institute of Israel, GII; see Table 5).The three JHD clusters included one in the north with 14 events, a central one consisting of 12 events and one in the south with 16 events. There was also some overlap among the three JHD clusters, so that a few events in the northern and southern clusters were common to a few events in the central cluster. The JHD results were constrained by one event for each cluster using GT5 HDC solutions of events promoted on the basis of their error ellipse sizes. Table 6 lists events common to HDC and the three JHD clusters. The listed source data are all HDC GT5 solutions, except for that of the reference event by GII.

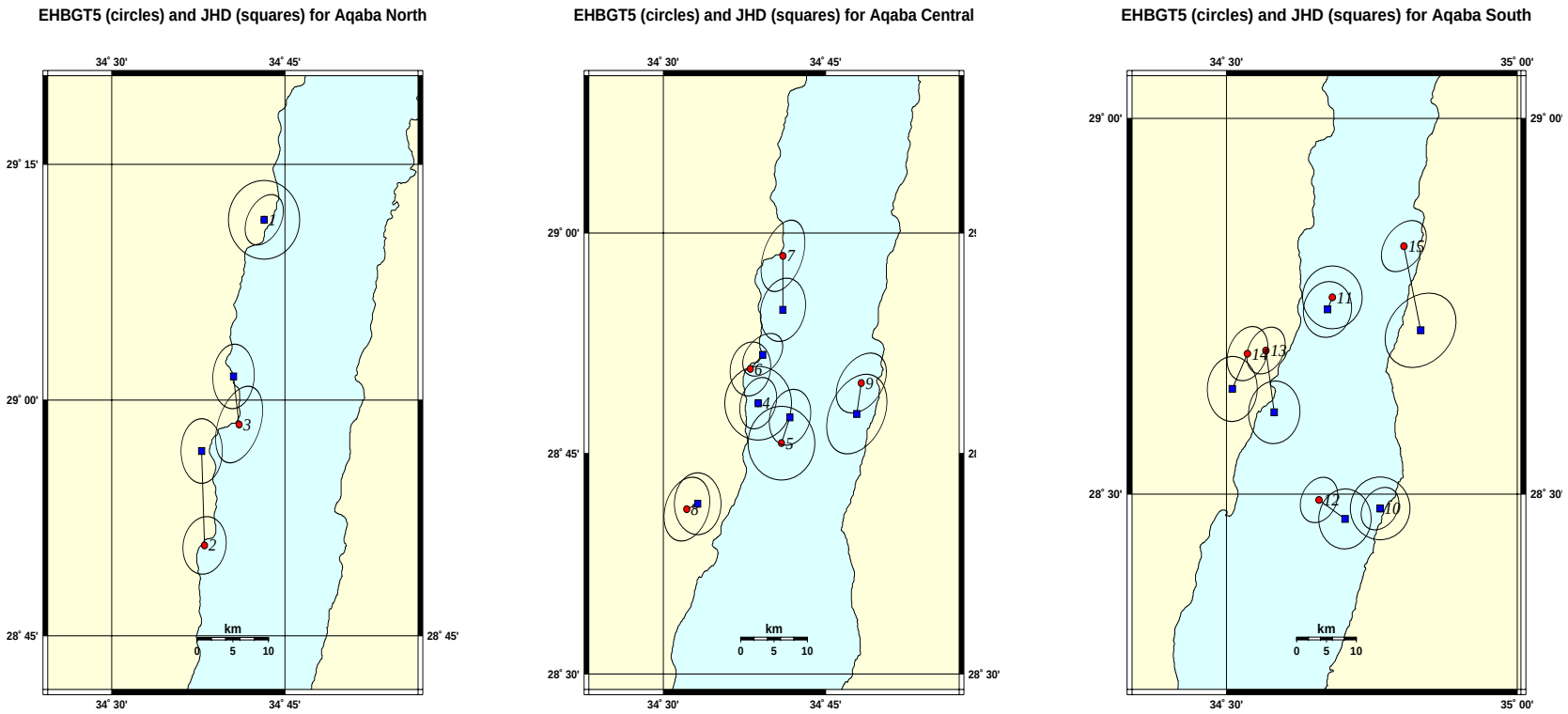

Figure 7. Comparison of JHD locations (squares) and GT5 reference events (circles) for the northern (left frame), central (middle frame) and southern Aqaba event clusters (right frame); event numbers refer to Table 5. The GII reference event (event no 5) and events constrained in the JHD solutions (event no 1, in northern cluster, event no 4 in central cluster, and event no 10 in southern cluster) are assigned an error circle with a 5 km radius, while estimated error ellipses are shown for HDC solutions of other events and for JHD locations.

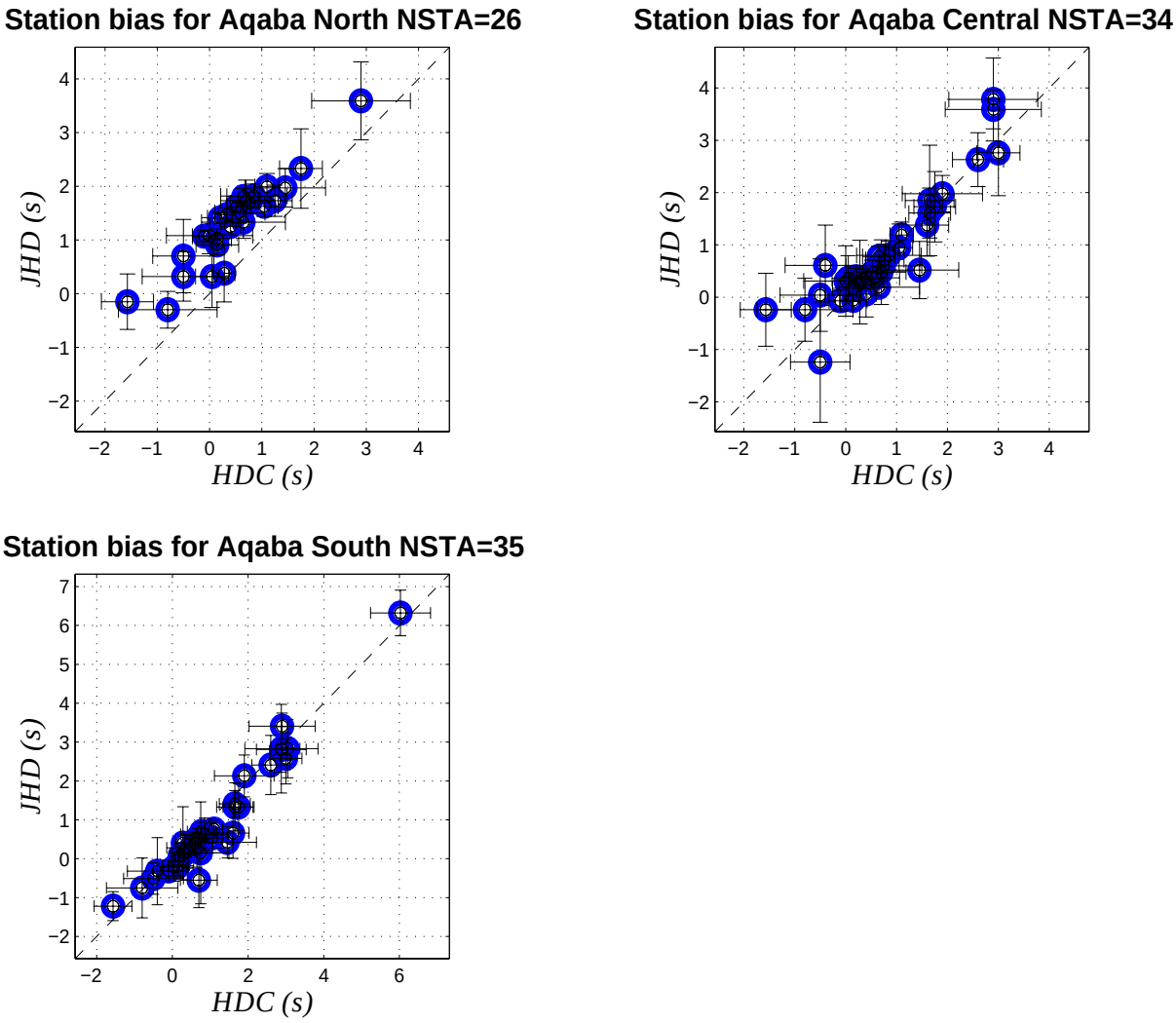

Figure 8. Travel time corrections for stations at epicentral distances < 20^{o} obtained by HDC and JHD for Aqaba northern (top left), central (top right), and southern clusters (bottom). Error bars indicate 90% confidence limits.

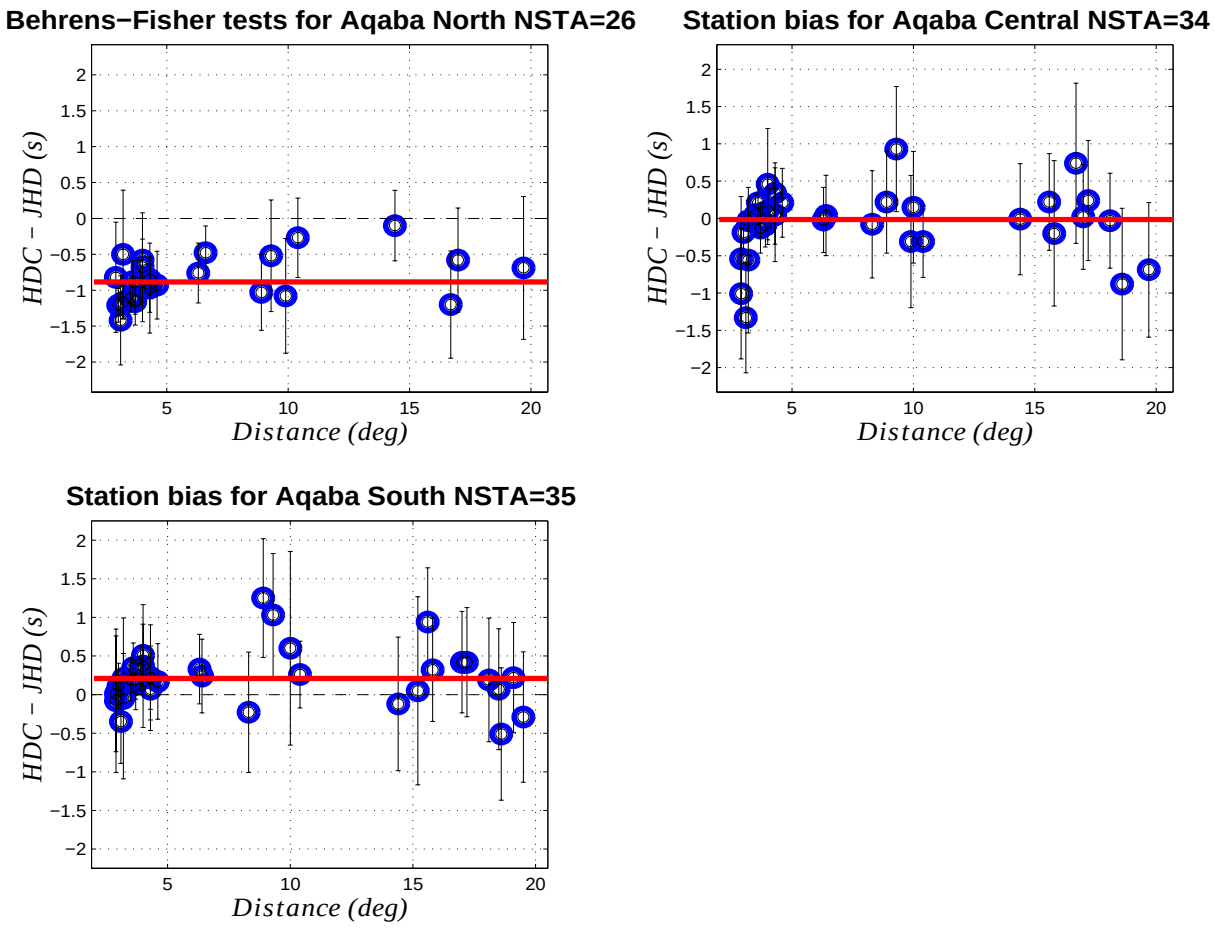

Figure 9. Difference in estimated travel corrections plotted as a function of epicentral distance for Aqaba northern (top left), central (top right) and southern event clusters (bottom). Error bars represent approximate 90% confidence limits (Behrens-Fisher) for the differences relative to the median of the differences (0.9 sec for northern, 0 sec for central, and 0.2 sec for southern Aqaba), which are marked as a thick line in each frame.

The maps in Figure 7 compare HDC and JHD locations. Event held fixed for the JHD locations (event nos. 1, 4, and 10) and the GII reference event (event no. 5) are assigned error circles with a 5 km radius. Other error ellipses represent HDC and JHD estimates, which all overlap for the central cluster. There is no overlap for event no 2 in the northern cluster and the events 13 and 15 in the southern cluster. The constraining events are only in the category GT5, but if they were shifted a few km (event no in the northern cluster shifted to the south and event no 10 in the southern cluster shifted to the north) overlap would be achieved. That is to say that the lack of overlap for the three events in the southern and northern clusters is still consistent with the assumed and estimated uncertainties. However, the consistency in HDC and JHD event solutions for the Aqaba clusters is not as pronounced as for the Racha clusters discussed above.

Table 5: GT5 solutions of Aqaba reference events common to HDC and JHD clusters

Event no	Region	Date	Time	Lat(N)	Lon(E)
1	North[a]	1995/11/23	18:07:16.7	29.1910	34.7200
2	North	1995/12/11	01:32:06.9	28.8460	34.6340
3	North	1995/11/24	16:43:46.5	28.9740	34.6840
4	Central[a]	1996/02/21	04:59:52.1	28.8070	34.6460
5	Central[b]	1995/11/22	04:15:12.0	28.7620	34.6820
6	Central	1995/12/11	01:32:06.9	28.8460	34.6340
7	Central	1995/11/24	16:43:46.5	28.9740	34.6840
8	Central	1993/11/08	01:06:03.2	28.6870	34.5360
9	Central	1996/02/26	07:17:27.9	28.8300	34.8050
10	South[a]	1995/11/22	12:47:04.7	28.4810	34.7640
11	South	1995/11/22	04:15:12.0	28.7620	34.6820
12	South	1995/11/22	22:16:54.6	28.4920	34.6590
13	South	1993/11/03	18:39:33.2	28.6910	34.5680
14	South	1993/11/08	01:06:03.2	28.6870	34.5360
15	South	1996/02/26	07:17:27.9	28.8300	34.8050

a. Constrained in JHD
b. Reference Event by GII

Figure 8 compares travel time corrections for common stations at distances less than 20^{o}. Data are plotted only for stations with estimates based on 5 or more events for both HDC and JHD corrections. Furthermore the HDC corrections were based on residuals only for events with semi-major axes < 5.5 km. As for Azgir and Racha, there is general agreement between the two types of estimates, with an offset of about 0.9 sec for the northern Aqaba cluster. This offset can again be

attributed to differences in normalization of origin times. The comparisons in Figure 9 of the formal approximate confidence intervals for the differences in the estimated travel time corrections suggest differences on the 90% significance level for 3 out 26 stations for northern Aqaba, 3 out of 34 stations for central Aqaba and 1 out of 35 stations for southern Aqaba. Hence, the observed differences are reasonably consistent with the statistical variations due to differences in data and procedures.

CONCLUDING REMARKS

In the computational exercises above we have compared two approaches of event clusters analysis and their resulting estimates of epicenters and station travel time corrections. The two approaches, based on different and independent algorithms (HDC and JHD), were applied to identical data (Azgir) as well as independently compiled and only partly overlapping data (Racha and Aqaba). Estimated HDC and JHD locations and travel time corrections were in good agreement for all six tested cases. Differences in the estimates were generally consistent with their estimated uncertainties. The standard deviations of the differences in estimated travel time corrections ranged from about 0.1 sec (Azgir) to 0.35 sec for the Racha and Aqaba clusters. The areas of the error ellipses for Azgir events, which were processed with identical input data for HDC and JHD, were, on average, about 20% larger for the JHD estimates. The HDC ellipses are based on *a priori* assumptions about reading errors while the JHD estimates are based essentially on the *a posteriori* residuals. As a practical benefit of the computational exercises it can be mentioned that inconsistencies in co-ordinates and elevations of a few stations could be detected and corrected.

REFERENCES

Cox, D.R. and D.V. Hinkley (1974), Theoretical Statistics, Chapman & Hall.

Dewey, J.W. (1972), Seismicity and tectonics in western Venezuela, BSSA, 62: 1711-1751.

Dewey, J. W. (1989), JHD89 and SE89 - Algorithms for Joint Hypocenter Determination and modified Single-Event Determination (Memo USGS).

Douglas, A. (1967), Joint Epicenter Determination, Nature, 251:47-48.

Engdahl, E. R., van der Hilst, R., and R.; Buland (1998), Global Teleseismic Earthquake Relocation with Improved Travel Times and Procedures for Depth Determination, BSSA, 88: 722-743.

ISC (2001), http://www.isc.ac.uk

Jordan, T.H., Sverdrup, K.A. (1981), Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, BSSA, 71: 1105 - 1130.

NEIS (2001),http://neic.usgs.gov

Sultanov, D. D.; Murphy, J. R.; Rubenstein, K. D. (1999), A Seismic Source Summary for Soviet Peaceful Nuclear Explosions, BSSA 89: 640 - 647.